

Filière Génie Logiciel

Rapport de Projet de fin de deuxième année

Système de recommandation des hôtels

Réalisé par :

Aymane ANAM
Sabir AMHAIDA

Encadré par :

Mme Sanaa ELFKIHI

Année universitaire : 2019/2020

Remerciements

Au terme de ce travail, il nous est agréable d'exprimer notre reconnaissance auprès de toutes les personnes qui nous ayant aidé de près ou de loin à réaliser ce Projet. Tout d'abord nous adressons nos remerciements et notre gratitude à Madame ELFKIHI Sanaa, notre encadrante de PFA qui s'est montré généreuse en termes de temps et d'informations.

Nous ne manquons pas l'occasion de remercier chaleureusement tout le corps professoral et le cadre administratif, qui incarnent les valeurs du travail, qualité et ouverture au sein de l'Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes ENSIAS.

Cordialement.

Résumé

Notre projet de fin de la deuxième année consiste à réaliser une application WEB de recommandation des hôtels. Après une étude de l'existant et l'analyse des systèmes de recommandations, Nous avons pu définir un système hybride de recommandations dans lequel nous avons combiné entre trois approches, la première à base de modèles et dans laquelle nous avons comparé la précision de la prédiction du score donné par un usager à un hôtel, avec le modèle K-NN, Random forest et XGBoost. la deuxième approche est basée sur le filtrage collaboratif basé sur l'utilisateur et son activité sur la plateforme, et la dernière est basée sur le contenu des hôtels sans prendre en considération les activités des utilisateurs.

Mots-clés : Système de recommandation, K-NN, XGBoost, Random Forest, filtrage collaboratif, filtrage basé sur le contenu

Abstract

Our final year project consists in creating a WEB application for hotels recommendations. We were able to make a hybrid recommendation system in which we combined between three approaches, the first based on models and in which we compared the accuracy of the prediction of the score given by a user to a hotel, with the K-NN model, Random forest and XGBoost, the second is based on collaborative filtering based on the user and his activity on the platform, and the last is based on the content of the hotels without taking into account the activity of the users.

KeyWords : Recommendation system, hybrid system, K-NN, XGBoost, Random Forest, collaborative filtering, content-based filtering

Table des matières

Table des figures	8
1 Introduction générale	1
1.1 Contexte du projet	1
1.2 Objectifs	1
1.3 Exploration des solution existantes	2
1.4 Organisation du projet	2
2 Généralités sur les Systèmes de recommandations	3
2.1 Introduction	3
2.2 Définition	3
2.3 Intérêts et domaines d'application	4
2.3.1 Intérêts	4
2.3.2 Exemples de domaines d'applications	4
2.4 Classification	5
2.4.1 Système de recommandation basé sur un modèle	5
2.4.2 Système de recommandation basé sur une heuristique	5
2.5 Conclusion	8
3 Implémentation des algorithmes de recommandations proposés	9
3.1 Outils utilisés	9
3.1.1 Outils utilisés pour la collection de données	9
3.1.2 Outils utilisés pour le travail sur les données	10
3.2 Data-set utilisée	11
3.3 Implémentation de modèles	14
3.3.1 Algorithmes à base de modèles	14
3.3.2 Algorithmes basés sur la mémoire (Modèle du filtrage collaboratif à base d'utilisateur)	18

3.3.3	Implémentation d'un modèle basé sur le contenu des hôtels	18
3.4	Scénarios (Diagrammes BPMN)	21
3.4.1	Definition du BPMN	21
3.4.2	BPMN de notre système de recommandation proposé	21
3.5	Conclusion	23
4	Réalisation de l'application Web	24
4.1	Outils utilisés	24
4.2	Captures d'écran	26
4.2.1	Authentification	26
4.2.2	Section Sign in	27
4.2.3	Section liste des hôtels	27
4.2.4	Page des hôtels	30
4.3	Conclusion	32
5	Conclusion et perspectives	33
Bibliographie		34

Table des figures

2.1	Filtrage collaboratif [1]	6
2.2	Content-based recommender systems [1]	8
3.1	Colaboratory logo	10
3.2	Numpy logo	10
3.3	Pandas logo	11
3.4	scikit learn logo	11
3.5	Illustration du site TripAdvisor	12
3.6	Capture d'écran du fichier csv construit après les scraping	13
3.7	Capture d'écran du fichier csv des scores donnés par les utilisateurs	13
3.8	Corrélation entre le score donné par les utilisateurs à un hôtel et les autres dimensions.	14
3.9	Modélisation du fonctionnement de l'algorithme K-NN	16
3.10	Formule de calcul de la MAPE (Mean absolute percentage error)	16
3.11	Formule de calcul du MSE	17
3.12	Tableau des Métriques calculées pour l'évaluation de certains modèles afin de prédire le score qui sera donné aux hôtels par l'utilisateur en question	17
3.13	Exemple d'un output du modèle basé sur le contenu	21
3.14	BPMN de notre projet	22
4.1	Bootstrap logo	24
4.2	Jakarta EE logo	25
4.3	MongoDb logo	25
4.4	Section de création de compte	26
4.5	Section de connexion	27
4.6	Recherche par nom	27
4.7	Recherche par prix	27
4.8	Liste de hôtels retournés	28

4.9	Page qui contient la liste de hôtels recherchés	29
4.10	Liste de hôtels retournés	29
4.11	Page « Hôtel page » qui contient les informations d'un hôtel donné	30
4.12	Informations détaillés des hôtels	30
4.13	Sous section de recommandation d'hôtels similaires situés dans la page « Hotel page»	31

Système de recommandation des hôtels

Chapitre 1

Introduction générale

1.1 Contexte du projet

Vu l'utilisation croissante des ordinateurs et des appareils mobiles et la tendance des gens à compter sur des services accessibles par internet, La digitalisation des services dans plusieurs secteurs est devenue obligatoire. Parmi ces secteurs, On trouve celui du tourisme qui est devenu très dynamique dans ces dernières années avec le e-tourisme, Dans ce contexte la digitalisation du processus de la réservation des hôtels est devenue indispensable. Cette digitalisation oblige d'élaborer des systèmes de recommandations intelligents et dynamiques qui prennent en considération l'activité des utilisateurs afin de proposer des recommandations de manière exacte et efficace des hôtels qui pourraient leurs intéresser, C'est dans ce contexte que se base notre projet, qui consiste a créer une plateforme de recommandations des hôtels.

1.2 Objectifs

L'objective du projet est de construire des modèles de recommandations des hôtels pour notre plateforme, qui est une application web de recommandation des hôtels. Le système de recommandations proposé doit assurer au maximum possible l'adéquation des résultats recommandés avec le profil et l'activité des utilisateurs, en plus de proposer plusieurs alternatives au cas où l'activité de l'utilisateur est insuffisante pour faire une prédiction. Notre projet donc ne consiste pas seulement à élaborer un modèle de recommandation mais aussi à prendre en considération tous les scénarios possibles en tenant compte des données de l'utilisateur et des hôtels disponibles.

1.3 Exploration des solution existantes

Il existe plusieurs plateformes de tourisme qui implémentent des systèmes de recommandation et qui offrent à leurs utilisateurs une facilité de recherche et des recommandations pertinentes, à titre d'exemples nous citons TripAdvisor et Expedia.

TripAdvisor

C'est un site web qui offre des avis et des conseils touristiques émanant de consommateurs sur des hôtels, restaurants, villes et régions, lieux de loisirs, etc., à l'international. Il fournit également des outils de réservation de logements et de billets d'avion comparant des centaines de sites web afin de trouver les meilleurs prix. [2]

Expedia

Expedia est une société exploitant plusieurs agences de voyages en ligne incluant Expedia.com, Hotels.com, Hotwire.com, Egencia (anciennement Expedia Corporate Travel), Venere, Expedia Local Expert, Classic Vacations et eLong. Expedia exploite plus de 90 marques de points de vente dans plus de 60 pays. Elle opère également des réservations de voyage pour un certain nombre de sociétés aériennes et d'hôtels, de marques grand public, de sites web à fort trafic, et de milliers d'affiliés actifs par le biais d'Expedia Affiliate Network. La société fait partie des deux indices boursiers S&P 500.

1.4 Organisation du projet

Ce projet donc a pour but d'implémenter un système de recommandation pour notre application Web de recommandation des hôtels. Pour ceci Nous devons tout d'abord savoir c'est quoi un système de recommandation et quelles sont les différentes approches pour l'implémenter, voici le plan qu'on va suivre dans notre projet :

- Chapitre 2 : dans ce chapitre, on va faire une exploration théorique du sujet
- Chapitre 3 : Dans ce chapitre. on va détailler les différents algorithmes de recommandation utilisés
- Chapitre 4 : Présentation de l'application web réalisées implémentant les systèmes de recommandation construits

Chapitre 2

Généralités sur les Systèmes de recommandations

2.1 Introduction

Chaque jour, dans les réseaux sociaux ou autre nous nous trouvons devant beaucoup d'options et de choix. Quel news ou livre lire ? Quelle musique écouter ou vidéo visionner ? Quel produit acheter ? Quel film regarder ? et la taille de ces listes de choix sont généralement très grande. Amazon a plus de 562 millions de produits dans son magasin [3] et Netflix qui compte 167 millions d'abonnés dans le monde [4] offre un grand nombre de films et de séries, le nombre de films regardés sur Netflix par un utilisateur moyen de la plateforme est 60 entre le 1er novembre 2016 et la même date en 2017 [5]. A partir de ces chiffre on peut dire que diriger ou bien aider l'utilisateur à découvrir et à choisir des items (produits ou services) dans les différentes plateformes et réseaux sociaux est un défi important qui reste toujours d'actualité.

Dans cette partie nous allons présenter l'aspect théorique de ce sujet y compris les définitions et les concepts de base des systèmes de recommandations et les techniques qu'ils utilisent ainsi que leurs limites.

2.2 Définition

Un système de recommandation est défini de plusieurs manières mais nous allons citer la définition la plus populaire qui est celle de Robin Burke [6] : **un Système de recommandation** : *Système capable de fournir des recommandations personnalisées ou permettant de*

guider l'utilisateur vers des ressources intéressantes ou utiles au sein d'un espace de données important. Donc d'après cette définition le rôle principale d'un système de recommandation est de prédire avec précision les items qui pourraient intéresser l'utilisateur.

2.3 Intérêts et domaines d'application

2.3.1 Intérêts

Pour l'utilisateur

Il y a plusieurs intérêts, à savoir la réduction du temps de recherche pour l'utilisateur, la découverte des produits difficiles à trouver et aussi recevoir des suggestions de la part du système auxquelles l'utilisateur n'aurait pas spontanément prêtées attention.

Pour fournisseur du service de systèmes de recommandations

Un système de recommandation est très utile pour le fournisseur, il lui permet par exemple de :

- Garder et fidéliser les clients.
- Orienter les clients.
- Augmenter les bénéfices.
- Fournir à un utilisateur des ressources pertinentes en fonction de ses préférences.

2.3.2 Exemples de domaines d'applications

Dès leurs apparitions, plusieurs systèmes de recommandations ont pu être bénéfiques dans nombreux domaines à savoir :

- Musique (deezer, ...)
- Jeux vidéo (Microsoft Xbox Live, ...)
- Livres (Amazon, ...)
- Films (Netflix, ...)
- Amis (LinkedIn, Facebook, Twitter, Google+, ...)
- Publicité(Facebook, Google Ads)
- e-Shopping (eBay, Amazon,...)
- etc.

2.4 Classification

Il existe plusieurs classifications des système de recommandation qui sont utilisés chacun dans un contexte bien particulier, mais on va s'intéresser aux celles les plus connues et sur lesquelles nous basons notre étude à savoir : [7]

2.4.1 Système de recommandation basé sur un modèle

Ce type de modèle repose essentiellement sur le fait d'entraîner un modèle afin de générer une recommandation sans utiliser directement l'historique des utilisateurs pour donner des recommandations.

2.4.2 Système de recommandation basé sur une heuristique

Aussi appelé système de recommandation basé sur la mémoire, Ce type de méthode consiste à utiliser les évaluations des usagers à chaque fois qu'une recommandation est calculée, et ne repose pas sur les calculs qui ont été précédemment effectués, C'est la raison principale de son surnom de modèle basé sur la mémoire, puisque le traitement est fait à chaque fois qu'une recommandation doit être faite. [8]

2.4.2.A Filtrage collaboratif (collaborative filtering)

Il s'agit du moyen le plus performant pour construire un système de recommandation qui est utilisé principalement dans le cas où les recommandations qu'on désire effectuer visent des utilisateurs déjà connus, ou des personnes ayant déjà utilisés nos services, Ce type de système de recommandations repose sur un principe nommé **Adage**, qui stipule que deux personnes ayant appréciés un contenu dans le passé sont forts probables d'être intéressés une nouvelle fois par le même contenu, Le filtrage collaboratif peut être classifié en trois types principaux qui sont le User-Based filtering, l'Item-based filtering et la factorisation de matrice. [9]

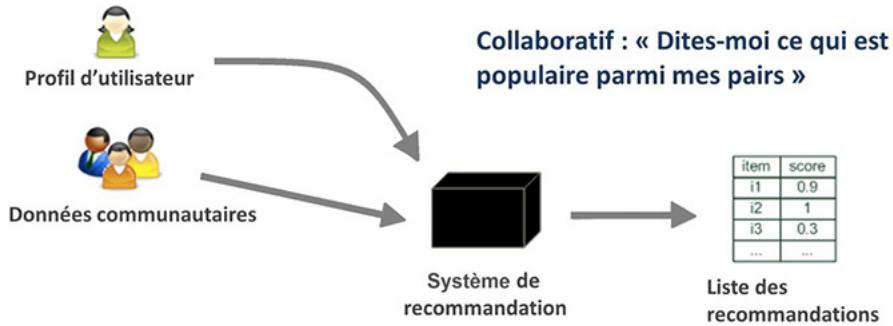


FIGURE 2.1 Filtrage collaboratif [1]

- **User-Based filtering (filtrage collaboratif basé sur l'utilisateur)**

Son principe repose essentiellement sur la comparaison entre les profils d'utilisateurs, en proposant un contenu qui est similaire à un contenu aimé par un autre utilisateur qui a le même profil. L'avantage principal de cette approche est qu'elle est agnostique et indifférente au contenu ce qui pourrait être très bénéfique dans certains domaines d'application des systèmes de recommandations. Exemples de cas d'application : la presse en ligne.

- **Item-Based filtering (filtrage collaboratif basé sur l'item)**

Ce type d'algorithme est un peu similaire au User-Based filtering, mais cette fois les similarités étudiées sont entre les items (articles, musiques, vidéos, fonctionnalités, etc) et non pas entre les utilisateurs ayant interagi avec ces items. Cette méthode utilise le fait que si deux utilisateurs A et B aiment un item X, alors un item Y ayant été évalué positivement par l'utilisateur A pourrait plaire à l'utilisateur B. Ce type d'algorithme commence le calcul en utilisant les contenus.

- **Factorisation de matrice (FC)**

C'est une méthode qui modélise le fait qu'un utilisateur soit intéressé par un item à l'aide d'une matrice, où la ligne i représente le user i et la colonne j représente l'item, la valeur (i,j) étant le score ou bien la note attribuée par l'utilisateur i à l'item j . Cette méthode essaie de prédire un score inconnu représenté généralement par le symbole ?, et cela en décomposant la matrice de base en des matrices plus petites. Le point faible qu'on peut rapprocher à cette méthode est qu'elle n'est pas efficace lorsqu'on est dans un contexte dynamique, où de nouveaux utilisateurs et items arrivent sans cesse. [10]

- **Systèmes de recommandation à base de confiance (SRC))**

Appelée aussi recommandation sociale, cette approche ressemble au système de recommandation FC basé sur l'utilisateur. L'une des différences entre ces deux types

de systèmes, et que le SRC repose sur la notion de confiance et non sur la notion de voisinage qui est basée sur la similarité entre utilisateurs, Ainsi l'utilisateur a la possibilité de choisir ses voisins qui sont nommés dans ce type de systèmes voisins. Un autre point de différence entre les deux systèmes est la notion de propagation de confiance, qui permet à un utilisateur de non seulement profiter des recommandations à partir d'un voisin, mais aussi à partir du voisin de son voisin.

Limitation du FC

- **Démarrage à froid :** Le FC souffre du problème du démarrage à froid dans ses deux types, celui qui est basé sur l'utilisateur, et celui qui est basé sur les items, Dans ce type de systèmes de recommandation, on ne peut pas donner des recommandations aux utilisateurs n'ayant pas encore évalués des items.
- **Résistance :** Les systèmes de recommandations FC sont vulnérables aux attaques malveillantes basées sur le retour des utilisateurs, puisqu'un utilisateur peut influencer les recommandations de ces voisins sans autorisation.
- **Passage à l'échelle :** Ce type de système est fragile au passage à l'échelle, puisqu'il demande des opérations de calcul de similarités assez importantes.

2.4.2.B Système de recommandation à base de contenu (Content-based recommender systems)

Cette approche est différente du filtrage collaboratif, et cela en traitant le contenu des items sans prendre en considération les utilisateurs ayant interagis avec, Ce type d'algorithme analyse le contenu des différents items se trouvant dans le système, et essaie de trouver des similarités, L'idée générale de ce type de système est de définir un thème principal concernant un item, L'un des cas d'utilisations de ce type de système est la recommandation des articles scientifiques.

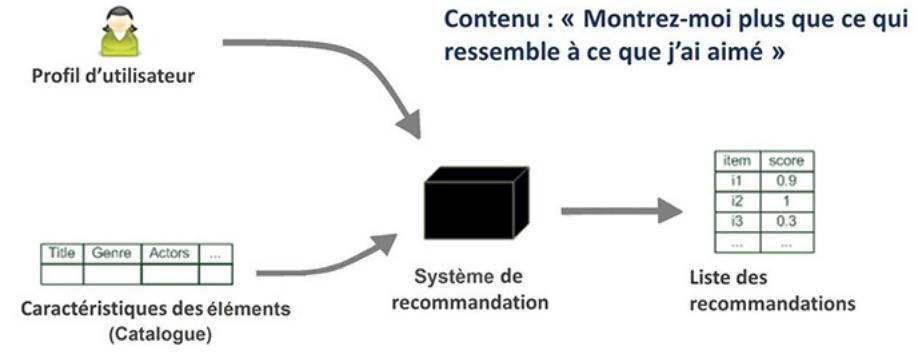


FIGURE 2.2 Content-based recommender systems [1]

Limitation du CB

- **Manque de données** : Ce type de système est fragile face au manque de données, puisque il requiert une description minutieuse des items afin de faire des recommandations.
- **Gestion de relation** : Ce type de système est incapable de trouver des relations entre des items qui peuvent intéresser simultanément des utilisateurs, mais dont le contenu est tout à fait différent.
- **Difficulté de conception** : Comme cité auparavant, ce type de système requiert des descriptions détaillées du contenu.

2.5 Conclusion

Dans ce chapitre nous avons mis l'accent sur les différentes parties théorique de notre projet à savoir la définition, et plusieurs approches des systèmes de recommandations. Dans le troisième chapitre nous allons essayer d'illustrer et implémenter ces algorithmes afin de implémenter un système hybride de recommandation qui combine plusieurs méthodes.

Chapitre 3

Implémentation des algorithmes de recommandations proposés

Ce chapitre est dédié pour l'implémentation des algorithmes de recommandations.

3.1 Outils utilisés

3.1.1 Outils utilisés pour la collection de données

Le web scraping est la récupération de données de pages web, de façons automatique. Le web scraping se fait en deux étapes :

- *Le téléchargement* du code HTML et pour cela nous avons utilisé la bibliothèque **requests** de Python et nous avons automatisé la navigation par **Selenium WebDriver**
- *Le parsing* Nous avons utilisé la bibliothèque **Beautiful Soup** de Python

BeautifulSoup4

Beautiful Soup est une bibliothèque **Python** d'analyse syntaxique de documents HTML et XML. Elle produit un arbre syntaxique qui peut être utilisé pour chercher des éléments ou les modifier. Elle est très connue dans le contexte de la collecte de données.

Selenium WebDriver

Selenium WebDriver est un framework open-source gratuit qui fournit une interface de programmation d'application (API) commune pour l'automatisation du navigateur. Idéalement, les navigateurs Web modernes devraient tous rendre une application Web de la même manière. Cependant, chaque navigateur possède son propre moteur de rendu et gère

le HTML un peu différemment. les tests automatisés qui utilisent l’API client Selenium peuvent s’exécuter sur n’importe quel navigateur avec un pilote de réclamation WebDriver, y compris Chrome, Safari, Internet Explorer, Microsoft Edge et Firefox. Selenium WebDriver peut fonctionner sur les plateformes Windows, Linux et macOS. Des tests peuvent être écrits pour Selenium WebDriver dans toutes les langage de programmation pris en charge par le projet Selenium, y compris Java, C#, Ruby, Python et JavaScript. Ces langages de programmation communiquent avec Selenium WebDriver en appelant des méthodes dans l’API client Selenium.

3.1.2 Outils utilisés pour le travail sur les données

Google Colab



FIGURE 3.1 Colaboratory logo

C’est un service cloud, offert gratuitement par Google, basé sur Jupyter Notebook. Cette plateforme permet d’entraîner des modèles de Machine Learning directement dans le cloud. Sans donc avoir besoin d’installer quoi que ce soit sur notre ordinateur à l’exception d’un navigateur.

Numpy

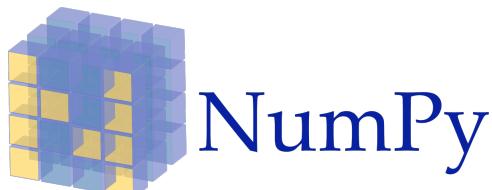


FIGURE 3.2 Numpy logo

NumPy est une extension du langage de programmation Python, elle contient les outils nécessaire pour la manipulation des matrices ou tableaux multidimensionnels ainsi que des fonctions mathématiques opérant sur ces tableaux, ce qui va nous aidez dans notre Projet.[11]

Pandas

Pandas est un outil d'analyse et de manipulation de données open source, rapide, puissant, flexible et facile à utiliser, construit sur le langage de programmation Python. Nous avons choisi à l'utiliser grâce à sa richesse et sa puissance dans l'analyse et la manipulation de données.[12]



FIGURE 3.3 Pandas logo

Scikit learn

C'est une bibliothèque Python de machine learning. elle propose divers algorithmes de classification, de régression et de clustering, y compris des machines à vecteurs de support, Random Forests, gradient boosting, k-means et DBSCAN, et est conçu pour interagir avec les bibliothèques numériques et scientifiques de Python comme NumPy et SciPy.



FIGURE 3.4 scikit learn logo

3.2 Data-set utilisée

Lors du déploiement de notre application, les données seront essentiellement extraites de la base de données, mais afin de choisir le modèle adéquat qui sera utilisé selon chaque cas d'étude, nous avons utilisé une dataset que nous avons récupéré par scraping en utilisant la bibliothèque de scraping BeautifulSoup, le site web utilisé pour extraire les données est <https://www.tripadvisor.com/> qui est un site de recommandation d'hôtel.

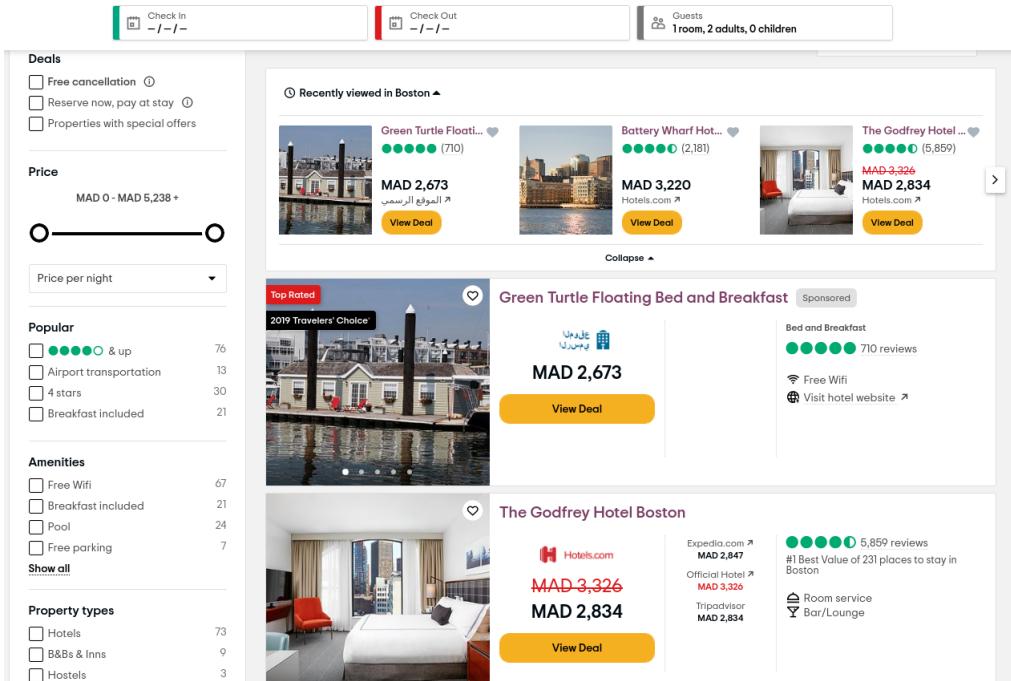


FIGURE 3.5 Illustration du site TripAdvisor

Les informations que nous avons extrait de chaque page d'hôtels sont les suivantes :

name : Nom de l'hôtel

url : Url de l'hôtel, qui sera utilisé pour extraire d'autres données si nécessaire.

Country : Le pays où se trouve l'hôtel.

Region : La région où se trouve l'hôtel.

location : Adresse exacte de l'hôtel.

property_amenties : Services proposés par l'hôtel (Exemples : Gym, piscine, parking, Wifi ..)

room_features : Propriétés des chambres de l'hôtel.

hotel_style : Style auquel l'hôtel appartient.

hotel_class : La classe de l'hôtel par exemple 5 étoiles, 3 étoiles ...

price : Le prix d'une nuit à l'hôtel.

hotel_score_reviews : La moyenne des scores attribués par les utilisateurs à cette hôtel entre 1 et 5.

totalReview : Nombre de reviews que l'hôtel a reçu.

location_score : Score associé à l'endroit où se situe l'hôtel entre 1 et 5.

cleanliness_score : Score associé à la propreté de l'hôtel entre 1 et 5.

service_score : Score associé au services proposés par l'hôtel entre 1 et 5.

value_score : Score associé à la valeur de l'hôtel entre 1 et 5.

name	Country	Region	location	property_amentie	room_features	hotel_style	hotel_ci	price	hotel_s	totalReview	location	cleanline	service_score	value_score
Radisson Blu	Sweden	Sweden	Stockho 1 Vasagatan P.O	Parking Free_Hi	Air_conditioning	Hi Business Central	4	1643	4	481	4.5	4.5	4	4
Hilton San Fra	United States	United States	Ca 333 O'Farrell Str	Valet_parking	Wi Air_conditioning	Hi Green Centrally_	4	3183	3.5	4483	4.5	4	4	3.5
Sofitel Londor	United Kingdom	United Kingdom	North Terminal G	Valet_parking	Fre Blackout_curtains	Business Central	2	2297	4.5	10401	5	4.5	4.5	4
Holiday Inn Lc	United Kingdom	United Kingdom	Gibson Road, Si	Parking Free_inte	Air_conditioning	Hi Centrally_Locate	3	1361.5	4	646	4	4	4	3.5
Radisson Blu	United Kingdom	United Kingdom	68-86 Cromwell I	Free_High_Spee	Air_conditioning	Hi Classic	4	2382	4	4945	4.5	4.5	4.5	4
Holmes Hotel	United Kingdom	United Kingdom	108 Baker Street	Free_High_Spee	Bathrobes	Air_con Classic Charming	4	3204	4.5	137	5	5	4.5	4.5
Hotel OTTO	Germany	Germany	Berlin (Kneisebeckstrass	Paid_private	par Allergy-free	room Family Mid-range	4	1236.5	4.5	1540	4.5	4.5	4.5	4.5
Bedruthan Ho	United Kingdom	United Kingdom	B3276 Trenance	Free_parking	Fre Private_beach	Fire Bay_View Great	4	2361.5	4.5	1614	5	4.5	4.5	4
Hotel du Vin E	United Kingdom	United Kingdom	Magdalen Street	Parking Free_Hi	Air_conditioning	Hi Trendy Quiet	4	1422	4	1806	4.5	4.5	4	4
Occidental Co	Mexico	Mexico Yucatan	I Boulevard Kukul Parkin	Wifi Pool	Air_conditioning	Pr Family Bay_View	4	2407	4	9416	4	4.5	4	4
Saiaz Getaria	Spain	Spain Basque C	Roke Deuna 25,	Parking Free_Hi	Air_conditioning	Pr Great_View	2	1030	4.5	408	5	4.5	4.5	4.5
Hotel Marque	Spain	Spain Basque C	Calle Torrea, 1, C	Free_parking	Fre Air_conditioning	Hi Quirky_Hotels	5	6174.5	4.5	1259	4.5	5	4.5	4
Imbali Safari	South Africa	South Africa Kru	Imbali Concessio	Free_parking	Fre Bathrobes	Air_con Quaint Great_Vie	5	7637.5	4.5	587	5	4.5	4.5	4.5
Franschhoek	South Africa	South Africa Wes	Main Road, Fran	Free_parking	Fre Bathrobes	Air_con Mountain_View C	5	2500	4.5	981	4.5	5	4.5	4.5
The Alexandre	South Africa	South Africa Wes	30 Glen View Ro	Free_parking	Fre Soundproof_rooms	Trendy Ocean_V	5	1498	5	468	5	5	5	5
Teremok Mari	South Africa	South Africa Kwa	49 Marine Drive,	Free_parking	Fre Air_conditioning	Hi Hidden_Gem Lu	5	2368.5	5	496	4.5	5	5	4.5
Dawsons Gan	South Africa	South Africa Mpum	Nelshoogte Area	Free_parking	Fre Room_service		5	1842	5	173	5	5	5	4.5
The Sands at	South Africa	South Africa Eas	8 Frank Road, S	Free_parking	Fre Air_conditioning	Pr Bay_View Quain	5	2313	4.5	253	5	5	4.5	4.5
Summerfields	South Africa	South Africa Mpum	R536 Sabie Roa	Free_parking	Fre Air_conditioning	Hi Luxury Great_Vie	5	3565	4.5	529	5	5	4.5	4.5
Villa Cora	Italy	Italy	Tuscani Prr. Viale Niccol	Ma Free_parkin	Fre Housekeeping	Priv Hidden_Gem Ro	5	5397.5	5	1307	4.5	5	5	4.5

FIGURE 3.6 Capture d'écran du fichier csv construit après les scraping

L'image ci-dessus contient les informations des hôtels collectées.

Nous avons aussi collectées les scores attribués au différents hôtels, Voici un exemple des données collectées.

username	hotel_name	score
Sarah Williams	Radisson Blu Royal Vik	4
Tomi Wickman	Radisson Blu Royal Vik	5
Howard M	Radisson Blu Royal Vik	5
Number34	Radisson Blu Royal Vik	5
pawelostrowski1	Radisson Blu Royal Vik	5
Carina A	Radisson Blu Waterfor	5
jarcall	Radisson Blu Waterfor	5
patricia r	Radisson Blu Waterfor	5
andrepannell	Radisson Blu Waterfor	5
Hüseyin K	Radisson Blu Waterfor	1
Sue G	Hilton Miami Airport Bl	5
Linds Helm	Hilton Miami Airport Bl	5
jennifer b	Hilton Miami Airport Bl	5
m0nchm0nch	Hilton Miami Airport Bl	3
infrequenttravell	Hilton Miami Airport Bl	4
Dennis E Smith	Crowne Plaza Hotel Sa	5
TravelinMama43	Crowne Plaza Hotel Sa	4
Nader H	Crowne Plaza Hotel Sa	3
M G	Crowne Plaza Hotel Sa	2
miesf	Crowne Plaza Hotel Sa	5

FIGURE 3.7 Capture d'écran du fichier csv des scores donnés par les utilisateurs

3.3 Implémentation de modèles

3.3.1 Algorithmes à base de modèles

Calcul de l'apport d'informations des différentes dimensions (Feature Significance)

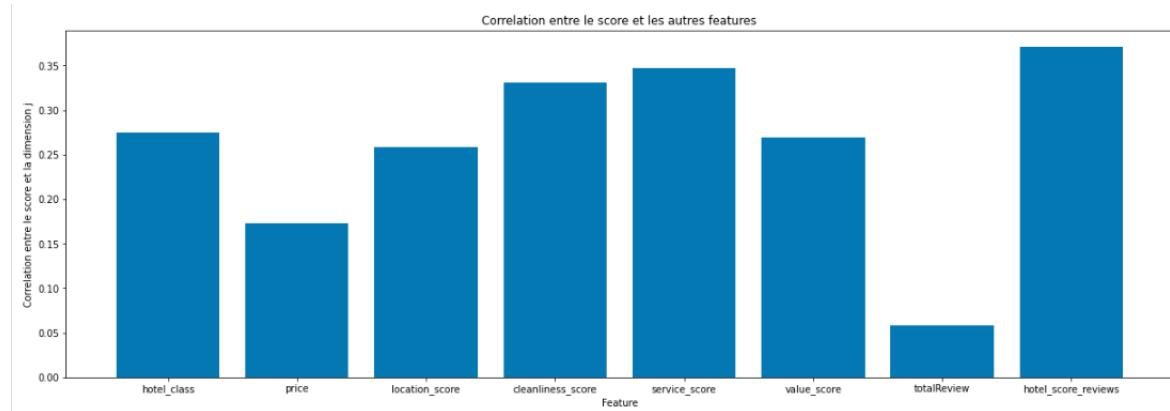


FIGURE 3.8 Corrélation entre le score donné par les utilisateurs à un hôtel et les autres dimensions.

On remarque que la dimension totalReview est presque indépendante de la dimension score ($\text{corr(score, totalReview)} < 6\%$), puisque la dimension score sera la dimension sur laquelle reposera notre modèle pour donner des recommandations, la dimension totalReview ne sera pas prise en compte à cause de son manque d'apport d'informations.

Préparation des données

En raison du problème de rareté des données et particulièrement le manque de reviews dans notre contexte qui peut se produire connue sous le nom de **Data sparsity**, notre traitement ne sera pas seulement appliqué aux hôtels ayant reçue un review de la part de l'utilisateur mais il sera également appliqué aux hôtels ayant reçu un review de la part des 40 plus proches voisins de l'utilisateur en question.

Différent modèles utilisés

Après l'extraction des 40 plus proche voisins, on prédit le score qui sera donné par l'utilisateur aux hôtels pour chacun des trois modèles qui sont K-NN, Random forests et XGBoost.

- **Présentation théorique du modèle XGBoost**

XGBoost(eXtreme Gradient Boosting) est un modèle d'apprentissage automatisé, qui repose sur le principe d'agrégation de modèles en utilisant des arbres de décision, et cela en combinant plusieurs modèles pour obtenir un seul output à la fin, l'apport du modèle XGBoost est comme mentionné précédemment, le modèle entraîne plusieurs modèles et fait un vote final pour choisir le bon output.

L'un des points forts de ce modèle est l'utilisation du principe de **Boosting**, qui essaie d'adapter le sous modèle courant aux erreurs commises par le modèle précédent, un autre point fort de ce modèle est l'utilisation de la régularisation afin d'éviter le problème du surajustement **Overfitting**.[13]

- **Présentation théorique du modèle Random forests**

C'est un modèle d'apprentissage automatisé appartenant à la famille des agrégations de modèle, ce modèle utilise le principe du **bagging**, qui est de donner des prévisions calculées à l'aide de plusieurs modèles. Si on est dans le cadre d'une classification la catégorie la plus proche est celle qui sera choisie, sinon dans le cas d'une régression on calculera la valeur moyenne.[14]

- **Présentation théorique du modèle K-NN**

Le modèle d'apprentissage automatisé utilisé dans cette approche est le modèle K-NN connue sous le nom de méthode des k plus proches voisins. La méthode consiste à prendre n entrées de dimensions p, et prédire une dimension i d'une entrée x en se basant sur les k plus proches voisins des n résultats retournés par le modèle. La valeur de la dimension i prédite est la classe la plus répétée dans les k plus proches voisins. Le calcul des plus proches voisins se fait en considérant chaque entrée comme un vecteur multidimensionnel et de rechercher les vecteurs les plus proches à notre entrée, en se basant sur une métrique, la métrique utilisée dans notre modèle est la distance cosinus.[15]

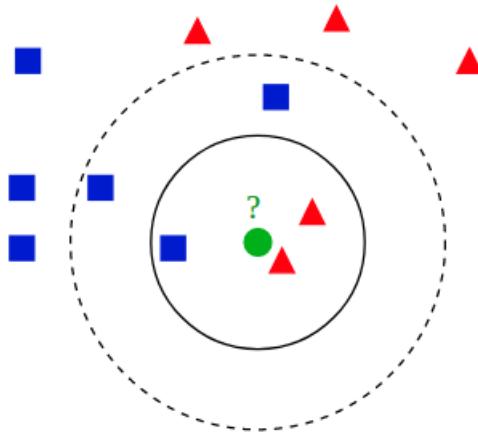


FIGURE 3.9 Modélisation du fonctionnement de l'algorithme K-NN

Métrique utilisée pour le choix du modèle

- **Performance personnalisé**

La mesure utilisée pour mesurer la performance est 1-MAPE. Si on note par y le score réel donné par l'utilisateur et \hat{y} le score prédict, la métrique MAPE sera définie comme suit :

$$MAPE = \frac{100\%}{n} \sum \left| \frac{y - \hat{y}}{y} \right|$$

FIGURE 3.10 Formule de calcul de la MAPE (Mean absolute percentage error)

- **F1 Score**

C'est la moyenne harmonique de la précision et du rappel. Elle prend en considération le résultat des faux positives et des faux négatives. La formule suivante permet de calculer Fscore : [16]

$$F = 2 \cdot \frac{(\text{précision} \cdot \text{rappel})}{(\text{précision} + \text{rappel})}$$

- La précision : décrit la proportion des items pertinents parmi l'ensemble des items proposés

- Le rappel : décrit la proportion des items pertinents proposés parmi l'ensemble des items pertinents
- **Le carré moyen des erreurs**

Connue sous le nom de MSE (Mean Square Error), c'est la moyenne arithmétique des carrés des écarts entre les valeurs prédictes par le modèle et les valeurs réelles, elle représente la variance résiduelle qu'on cherche à minimiser dans notre modèle. Notons Y le score réel, \hat{Y} le score prédict, et n le nombre totales des scores, la formule du MSE est donnée par

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

FIGURE 3.11 Formule de calcul du MSE

Estimation et choix du modèle

Le tableau de la figure 3.12 représente les valeurs obtenues des différentes métriques. On remarque que la performance et le F1 score sont plus élevés pour le modèle XGBoost. On remarque aussi que XGBoost a une MSE plus faible par rapport aux autres modèles. Ainsi, on choisit le modèle XGBoost pour prédire les scores qui seront attribués aux différents hôtels par l'utilisateur en question, et retourner la liste des hôtels qui seront recommandés par ordre décroissant du score prédict.

Nom du modèle	Performance	F1_score	mse
knn	0.745568	0.730102	1.35467
XGBoost	0.837418	0.811973	0.85742
RandomForest	0.830278	0.816808	0.899355

FIGURE 3.12 Tableau des Métriques calculées pour l'évaluation de certains modèles afin de prédire le score qui sera donné aux hôtels par l'utilisateur en question

3.3.2 Algorithmes basés sur la mémoire (Modèle du filtrage collaboratif à base d'utilisateur)

Préparation des données

Organisation de données sous forme de matrice où les lignes contiennent les utilisateurs et les colonnes contiennent la liste de tous les hôtels présents sur la plateforme. Pour un utilisateur i et un hôtel j , la valeur (i,j) vaut 1 si l'utilisateur a déjà consulté la page de l'hôtel et 0 sinon.

Modèle utilisé

On applique l'algorithme K-NN sur les données récemment construits, et on extrait la liste des utilisateurs les plus similaires. Le k utilisé est le nombre d'utilisateurs inscrits sur la plateforme.

Génération des recommandations

On retourne la liste des hôtels ayant le plus grand score donné par chaque utilisateur et cela par ordre décroissant de similarité entre l'utilisateur ayant effectué la requête et les autres utilisateurs de la plateforme. Les hôtels n'ayant pas été attribués un score par un utilisateur seront retournés par ordre décroissant de nombre de vues dans la plateforme et attachés aux résultats précédemment trouvés.

3.3.3 Implémentation d'un modèle basé sur le contenu des hôtels

Principe :

Pour ce modèle, la recommandation va se faire en se basant sur la similarité des hôtels. Cet algorithme analyse et exploite le contenu des différents hôtels indépendamment des interactions des usagers.

pour calculer la similarité, il existe plusieurs méthodes et parmi ces méthodes on trouve la similarité cosinus qui est la plus utilisée dans le contexte des systèmes de recommandations.

Similarité cosinus :

La similarité cosinus permet de mesurer la similarité entre deux vecteurs à n dimensions en déterminant le cosinus de l'angle entre eux[17]. si On note par X, Y deux vecteurs à n dimensions, nous pouvons formuler la similarité entre eux, s , par :

$$s(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} = \frac{\sum_{i=0}^{n-1} x_i y_i}{\sqrt{\sum_{i=0}^{n-1} (x_i)^2} \times \sqrt{\sum_{i=0}^{n-1} (y_i)^2}}$$

La valeur de $s(X, Y)$ est égale à $\cos \theta$ qui est entre -1 et +1, plus on s'approche de 1 plus l'angle θ diminué est donc la similarité augmente.

Préparation des données

Après le nettoyage de notre dataSet et pour adapter notre dataset avec notre approche, nous avons fait des modifications et des transformations sur notre dataset notamment sur les données qualitatives pour les rendre quantitatives. Pour ceci, nous avons utilisé la méthode **One-hot encoding**. Cette méthode répartit les valeurs d'une colonne sur plusieurs colonnes d'indicateurs et leur attribue 0 ou 1. Ces valeurs binaires expriment la relation entre la colonne groupée et la colonne codée.[18]

Nous avons utilisé cette méthode pour la colonne des biens immobiliers (property amenities) et aussi le style de l'hôtel (hotel style).

Pour bien illustrer cette méthode voilà un exemple dans notre contexte

Hotel_name	Property_amenities
Hilton San Francisco Union Square	Parking, Free_Wifi, Pool
Radisson Blu Royal Viking Hotel, Stockholm	Parking,Free_breakfast
Comfort Hotel Xpress Stockholm Central	Free_breakfast, Pool

One-Hot encoding

Hotel_name	Parking	pool	free_wifi	free_breakfast
Hilton San Francisco Union Square	1	1	1	0
Radisson Blu Royal Viking Hotel, Stockholm	1	0	0	1
Comfort Hotel Xpress Stockholm Central	0	1	0	1

TABLE 3.1 exemple de transformation de la colonne Property amenities

et de même pour la colonne hotel style et voici un petit exemple :

Hotel_name	hotel_style
Hilton San Francisco Union Square	Marina_View, Historic_Hotel, Classic
Radisson Blu Royal Viking Hotel, Stockholm	Classic, Historic_Hotel
Comfort Hotel Xpress Stockholm Central	Romantic

One-Hot encoding

Hotel_name	Marina_View	Historic_Hotel	Romantic	classic
Hilton San Francisco Union Square	1	1	0	1
Radisson Blu Royal Viking Hotel, Stockholm	0	1	0	1
Comfort Hotel Xpress Stockholm Central	0	0	1	0

TABLE 3.2 Exemple de transformation de la colonne hotel_style

Dans notre DataSet, nous avons pu extraire plus que 50 biens immobiliers et plus que 10 styles des hôtels.

Dimensions prises pour le calcul de la similarité

Pour calculer la similarité entre les hôtels, nous avons pris les dimensions suivantes :

- hotel_class : la classe de l'hôtel (entre 1 et 5)
 - hotel_price : moyenne de prix de réservation
 - hotel_score_reviews : le score de l'hôtel à partir des avis des usagers.
 - totalScoreReviews : nombre des avis de l'hôtel.
 - les différents biens immobiliers
 - les différents styles des hôtels

Après avoir centré ces dimensions, nous avons calculé la similarité entre elles par la méthode explicitée précédemment.

Calcul de similarité et visualisation des résultats

En se basant sur les données précédentes, nous obtenons les recommandations d'hôtels similaires, les différentes étapes suivies se résument comme suite :

- Calcul de la table de similarité.
 - La recherche d'indice de l'hôtel en input.
 - Extraction des k indices des hôtels de plus grands scores de similarité par rapport à notre hôtel.
 - Affichage des hôtels indexés par les indices extraits.

La figure 3.15 représente un exemple de input et l'output de ce modèle, dans cet exemple nous avons fixé le le nombre de recommandations en 10.

```
Input Hotel name: Radisson Blu Edwardian Vanderbilt
Recommendations for Radisson Blu Edwardian Vanderbilt:
1: Summerfields Rose Retreat & Spa
2: Hotel OTTO
3: Cape Heritage Hotel
4: Hotel Bedford
5: Hotel Mondschein
6: Park Hotel Tokyo
7: COSMO Hotel Berlin Mitte
8: Franschhoek Country House & Villas
9: Saiaz Getaria Hotel
10: London Marriott Hotel Regents Park
```

FIGURE 3.13 Exemple d'un output du modèle basé sur le contenu

3.4 Scénarios (Diagrammes BPMN)

3.4.1 Definition du BPMN

Le BPMN (Business Process Model and Notation) est une méthode de modélisation de processus d'affaires qui sert à décrire l'activité métier d'une organisation sous forme d'un modèle graphique, son but principale est de fournir un moyen qui soit lisible et compréhensible pour une grande tranche de personne dans un même organisme.[19]

3.4.2 BPMN de notre système de recommandation proposé

La figure 3.16 représente le schéma général de notre système de recommandation.

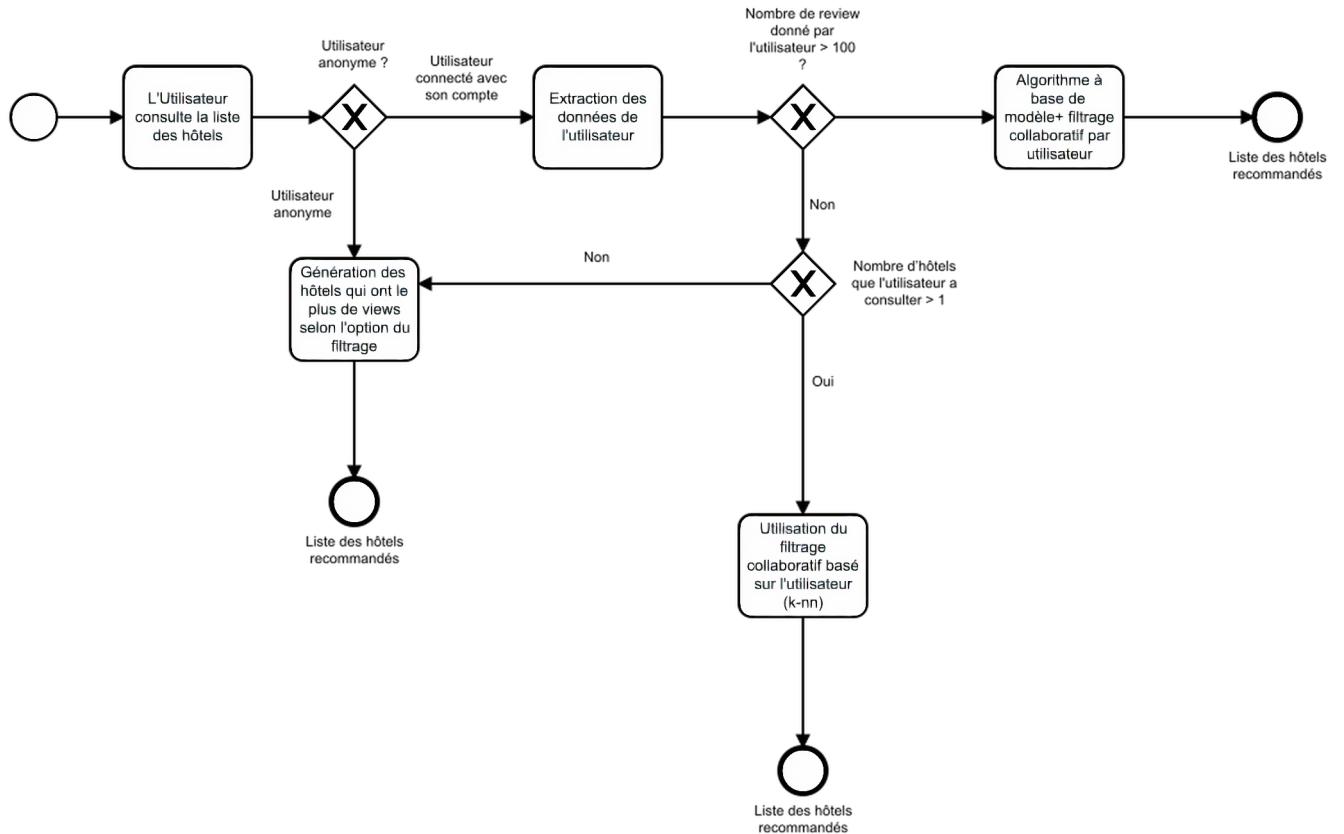


FIGURE 3.14 BPMN de notre projet

Cas d'utilisation de l'algorithme de filtrage collaboratif a base d'utilisateur

Cette recommandation est utilisée dans le cas où un utilisateur connecté sur un compte et ayant déjà visité une page d'un certain hôtel consulte la page liste des hôtels. La recommandation est aussi utilisée dans le cas du filtrage des hôtels par pays, classe ou par prix.

Dans le cas où un utilisateur ayant nouvellement rejoint la plateforme et n'ayant pas encore visité une page d'un hôtel, le problème du démarrage à froid est résolu en recommandant à l'utilisateur les hôtels ayant le plus de views dans la plateforme.

Cas d'utilisation de l'algorithme de recommandation à base de modèle

Cette recommandation est utilisée dans le cas où un utilisateur non anonyme ayant donné son review plus de 100 fois consulte la page liste des hôtels. La recommandation est aussi utilisé dans le cas du filtrage des hôtels par pays, classe ou par prix.

Cas d'utilisation de l'algorithme de recommandation à base de contenu des hôtels

Cette recommandation est utilisée dans le cas où l'utilisateur consulte la page d'un hôtel, on lui affiche des hôtels similaires à l'hôtel concerné.

3.5 Conclusion

Dans ce chapitre nous avons étudié plusieurs approches pour construire un système hybride de recommandation des hôtels. Le chapitre qui suit est consacré à la réalisation de l'application web qui implémente le système proposé.

Chapitre 4

Réalisation de l'application Web

Dans ce chapitre, nous allons tout d'abord présenter les outils de développement que nous avons utilisé pour réaliser notre application web, pour ensuite donner les interface de notre application qui implémente les différents algorithmes de recommandations explicités dans le chapitre précédent.

4.1 Outils utilisés

Bootstrap 4

Bootstrap 4 est un framework de front-end gratuit conçu pour développer la partie interface d'une application web plus rapidement et plus facilement. Il contient des modèles de conception HTML et CSS pour les formulaires, les boutons, les tableaux, la navigation, et bien d'autres, ainsi que des plugins JavaScript en option. Il donne aussi la possibilité de créer facilement des designs réactifs.[20]



FIGURE 4.1 Bootstap logo

Jakarta EE

Jakarta EE [21](anciennement J2EE, Java EE) est une spécification pour la plate-forme Java d'Oracle, destinée aux applications d'entreprise dans laquelle nous avons développé et exécuté notre application web en utilisant **IntelliJ IDEA Ultimate** comme environnement de travail.



JAKARTA EE

FIGURE 4.2 Jakarta EE logo

Hibernate OGM MongoDB

Hibernate OGM [22] est un moteur de persistance fournissant une prise en charge de la persistance Java (JPA) pour les DataStores de données NoSQL. Il réutilise le moteur de gestion du cycle de vie des objets d'Hibernate ORM, mais il conserve les entités dans un magasin NoSQL (clé / valeur, document, orienté colonne, etc.) au lieu d'une base de données relationnelle. Pour notre projet, nous avons travaillé avec MongoDB comme système de gestion de base de données NoSQL orienté documents.



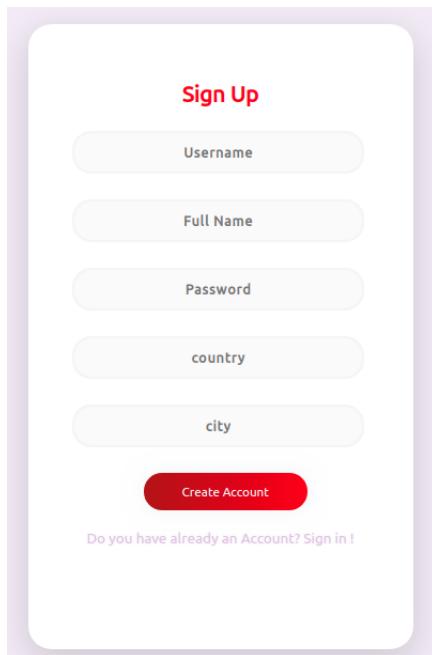
FIGURE 4.3 MongoDB logo

4.2 Captures d'écran

Afin d'assurer que la recommandation prenne en considération le comportement de l'utilisateur, nous avons intégré les fonctions d'authentification, pour que chaque utilisateur puisse disposer de son propre compte, et que ses préférences puissent être sauvegardés afin que la recommandation soit précise le plus possible.

4.2.1 Authentification

section Sign up



The image shows a 'Sign Up' form for account creation. The form is contained within a light gray rounded rectangle. At the top center, the text 'Sign Up' is displayed in red. Below this, there are five input fields, each with a placeholder label: 'Username', 'Full Name', 'Password', 'country', and 'city'. Each input field is enclosed in a light gray rounded rectangle. At the bottom of the form is a red rectangular button with the white text 'Create Account'. Below the button, a small line of text in gray reads 'Do you have already an Account? Sign in !'.

FIGURE 4.4 Section de création de compte

La figure 4.4 représente la page d'enregistrement d'un nouveau utilisateur.

4.2.2 Section Sign in

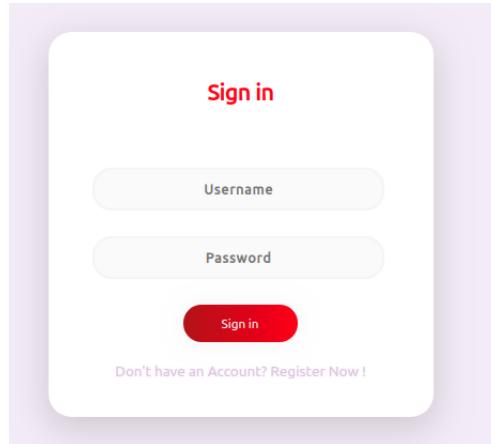


FIGURE 4.5 Section de connexion

La figure 4.4 représente la page de connexion de note application web.

4.2.3 Section liste des hôtels

Menu de recherche

Dans la section « Hotels list », l'utilisateur peut consulter la liste des hôtels présents sur la plateforme, les résultats retournés utilisent les différents modèles de recommandation mentionné ci-dessus selon le diagramme BPMN. Dans cette section, l'utilisateur a la possibilité de chercher un hôtel soit par nom, soit par pays, soit par classe ou par prix.

FIGURE 4.6 Recherche par nom

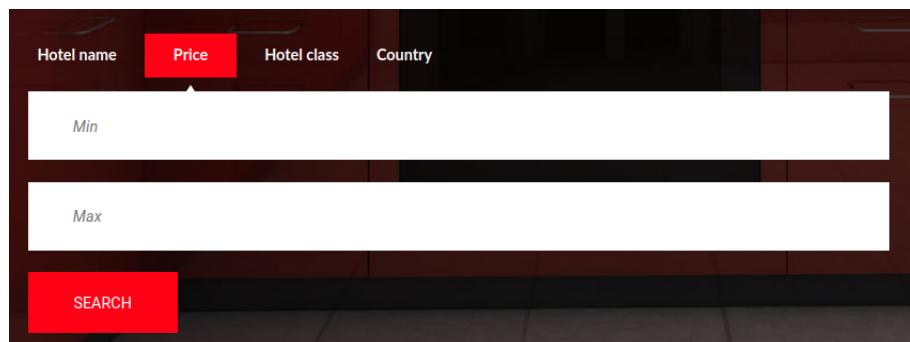


FIGURE 4.7 Recherche par prix

Lorsque la recherche se fait à l'aide d'un attribut quantitative (Price et le Hotel class), l'utilisateur a la possibilité de faire une recherche par intervalle.

Résultats retournés pour un utilisateur anonyme

Comme précisé auparavant dans le diagramme BPMN, lorsqu'un utilisateur navigue dans cette section, le résultat retourné contiendra la liste des hôtels qui ont le plus de views sur la plateforme.

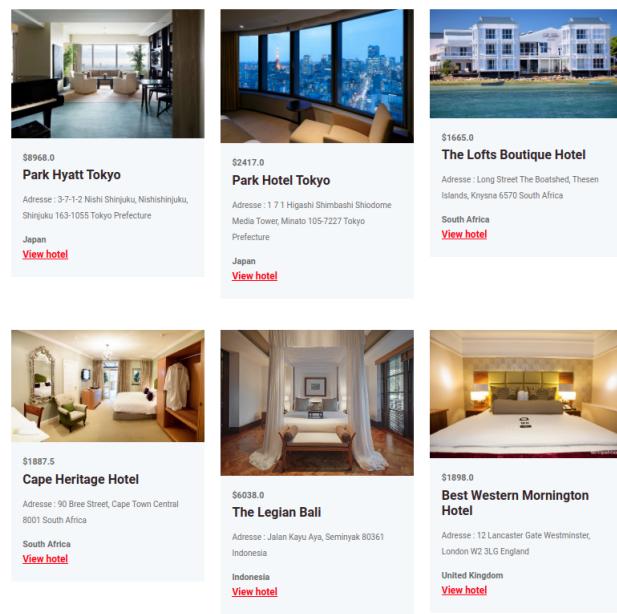


FIGURE 4.8 Liste de hôtels retournés

Résultats retournés pour un utilisateur inscrit sur la plateforme

Dans cette section, les résultats seront retournés en prenant en considération l'historique de l'activité de l'utilisateur. Le modèle utilisé pour la recommandation sera comme celui précisé dans le diagramme BPMN.

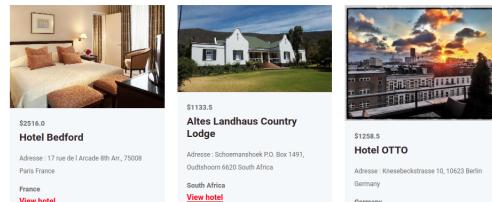


FIGURE 4.9 Page qui contient la liste de hôtels recherchés

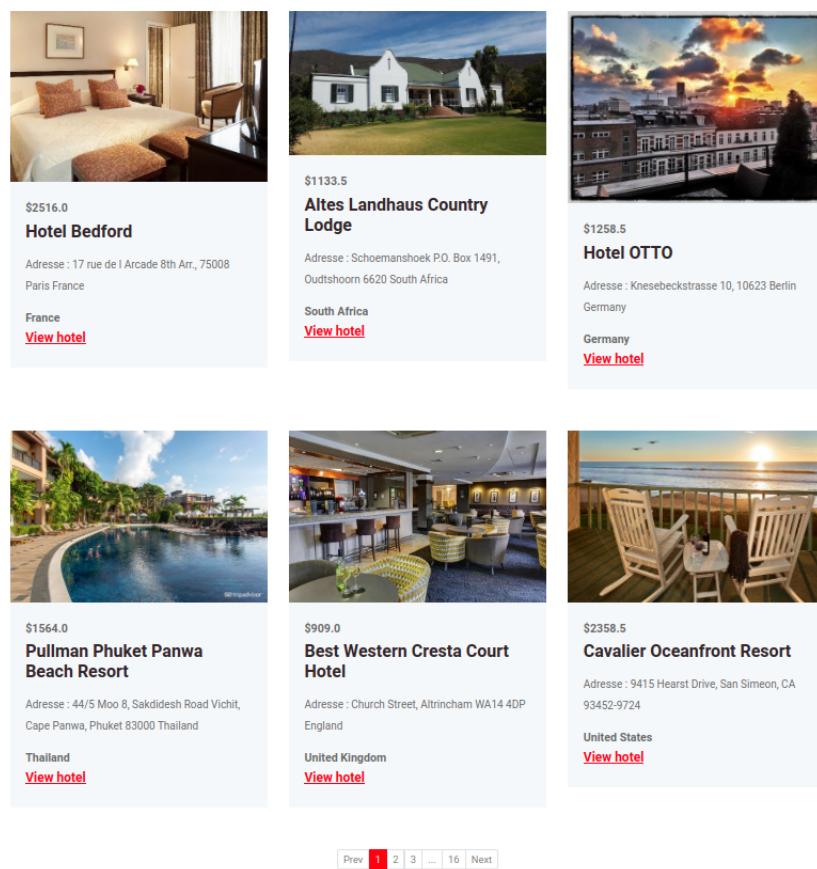
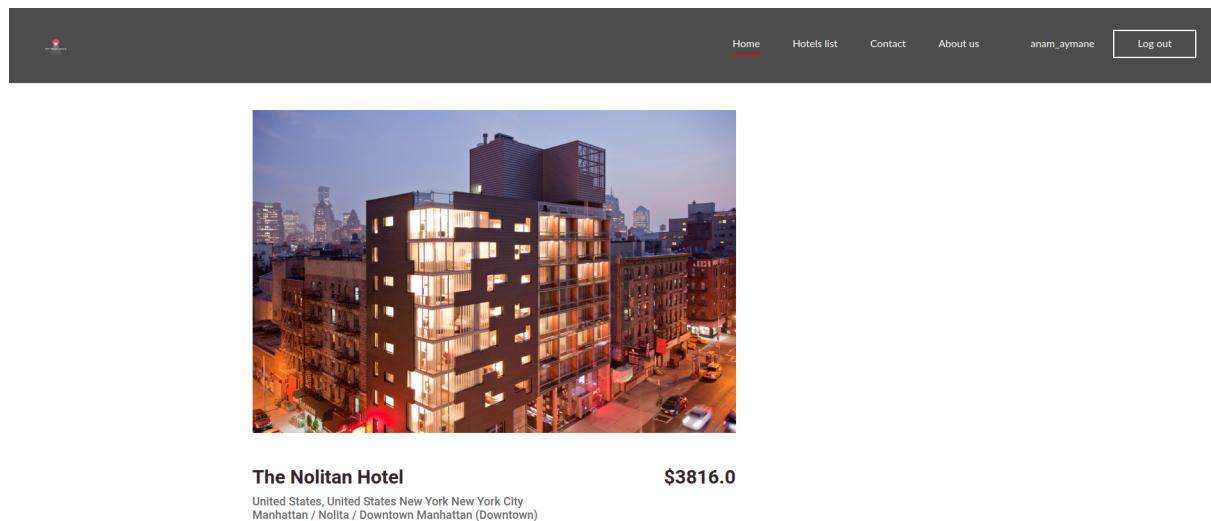


FIGURE 4.10 Liste de hôtels retournés

4.2.4 Page des hôtels

Cette page contient toutes les informations de l'hôtel, comme le prix, la classe de l'hôtel, son adresse, ses propriétés détaillés et d'autres informations. En bas de la page, les hôtels similaires à cet hôtel sont affichés par ordre de similarités suivant le modèle des plus proches voisins comme décrit auparavant.



The Nolitan Hotel **\$3816.0**
United States, United States New York New York City
Manhattan / Nolita / Downtown Manhattan (Downtown)

Property amenties	Room features	Hotel style
Valet_parking	Air_conditioning	Modern
Free_High_Speed_Internet_(WiFi)	Housekeeping	Romantic
Bar_lounge	Private_balcony	
Bicycle_rental	Safe	
Pets_Allowed_(Dog_Pet_Friendly)	Telephone	
Car_hire	Minibar	
Fax_photocopying	Flatscreen_TV	
Rooftop_terrace		

FIGURE 4.11 Page « Hôtel page » qui contient les informations d'un hôtel donné

Description				
Property amenties	Room features	Hotel style		
Valet_parking	Air_conditioning	Modern		
Free_High_Speed_Internet_(WiFi)	Housekeeping	Romantic		
Bar_lounge	Private_balcony			
Bicycle_rental	Safe			
Pets_Allowed_(Dog_Pet_Friendly)	Telephone			
Car_hire	Minibar			
Fax_photocopying	Flatscreen_TV			
Rooftop_terrace				

Detail				
Hotel score reviews	Location score	Cleanliness score	Service score	Value score
4.5	4.5	5.0	4.5	4.5

Give us your feedback for this hotel :

★★★☆☆

FIGURE 4.12 Informations détaillés des hôtels

Section qui contient des informations détaillées sur l'hôtel, en plus d'une zone où l'utilisateur peut donner son rating de l'hôtel.

Similar hotels

			
\$5611.0 The Bowery Hotel Hotel class : 5.0 United States Find out more	\$3987.5 Saybrook Point Inn & Spa Hotel class : 3.0 United States Find out more	\$6367.0 Andaz 5th Avenue Hotel class : 4.0 United States Find out more	\$6285.0 Limelight Hotel Hotel class : 4.0 United States Find out more
			
\$3183.0 Hilton San Francisco Union Square Hotel class : 4.0 United States Find out more	\$3157.5 Chamberlain West Hollywood Hotel class : 4.0 United States Find out more	\$3816.0 The Nolitan Hotel Hotel class : 4.0 United States Find out more	

FIGURE 4.13 Sous section de recommandation d'hôtels similaires situés dans la page « Hotel page»

Recommandation d'hôtel similaires

Comme cité auparavant, cette section contient les hôtels similaires à cet hôtel. Cette similarité a été calculé en utilisant un modèle de recommandation basé sur le contenu.

4.3 Conclusion

Dans ce chapitre nous avons vu l'implémentation des différents modèles de recommandations cités auparavant sous forme d'une application web.

Chapitre 5

Conclusion et perspectives

Le présent projet a pour sujet système de recommandation des hôtels était une occasion pour découvrir et travailler sur les systèmes de recommandation qui ont fait une révolution informatique dans le domaine des services web. Nous avons essayé d'implémenter un système hybride de recommandation des hôtels qui combine trois algorithmes, chacun dans un cas d'utilisation. Le premier algorithme est basé sur le contenu des hôtels, le deuxième est un système à base de modèle, et le dernier est un système basé sur le filtrage collaboratif à base d'utilisateur.

Notre perspectives est d'élargir notre DataSet et d'améliorer notre système pour augmenter la précision des recommandations, Nous souhaiterons aussi collecter d'autres activités de l'utilisateur comme sa position courante de l'utilisation, et aussi sa raison de voyage afin de recommander des résultats plus pertinents qui dépendent de son contexte.

Bibliographie

- [1] Les systèmes de recommandation : une catégorisation. URL <https://interstices.info/les-systemes-de-recommandation-categorisation/>.
- [2] Tripadvisor. [Wikipedia] <https://fr.wikipedia.org/wiki/TripAdvisor>.
- [3] Scaphero 2018. URL <https://www.scrapehero.com/number-of-products-on-amazon-april-2019/>.
- [4] Netflix compte 167 millions d'abonnés dans le monde. URL <https://www.20minutes.fr/high-tech/2702091-20200123-netflix-passe-le-cap-des-167-millions-d-abonnes>.
- [5] 140 millions d'heures d'écoute chaque jour sur netflix en 2017. URL <https://www.ledevoir.com/culture/ecrans/515210/140-millions-d-heures-ecoutees-sur-netflix-chaque-jour-en-2017>.
- [6] R Burke. *Hybrid recommender systems : Survey and experiments. User modeling and user-adapted interaction, 331–370.* 2002.
- [7] Idir Benouaret. Un système de recommandation contextuel et composite pour la visite personnalisée de sites culturels. *These*, pages 10–20, 2017. URL <https://tel.archives-ouvertes.fr/tel-01767997/document>.
- [8] Damien DUDOGNON. Diversité et système de recommandation :application à une plateforme de blogs à fort trafic. *These*, 2014. URL <http://thesesups.ups-tlse.fr/2546/1/2014TOU30203.pdf>.
- [9] Principaux algorithmes de recommandation. URL <https://www.mediego.com/fr/blog/principaux-algorithmes-de-recommandation>.
- [10] Naak Amine. Un système de gestion et de recommandation d'articles de recherche. *mémoire*, 2014. URL <https://core.ac.uk/download/pdf/55645772.pdf>.
- [11] numpy. [numpy] <https://fr.wikipedia.org/wiki/NumPy>.
- [12] pandas wikipedia. [pandas] <https://pandas.pydata.org/>.
- [13] La star des algorithmes de ml : Xgboost. [XGBoost] <https://www.datacorner.fr/xgboost/>.
- [14] Random forest, comment ça marche ? [Random Forest] <https://lovelyanalytics.com/2016/08/20/random-forest-comment-ca-marche/>.

- [15] Méthode des k plus proches voisins. [Wikipédia] https://fr.wikipedia.org/wiki/Méthode_des_k_plus_proches_voisins.
- [16] Précision et rappel. https://fr.wikipedia.org/wiki/Pr%C3%A9cision_et_rappel.
- [17] distance cosinus. [Wikipédia] https://fr.wikipedia.org/wiki/Similitude_cosinus.
- [18] Fundamental techniques of feature engineering for machine learning. [Towards Data Science] <https://towardsdatascience.com/feature-engineering-for-machine-learning-3a5e293a5114>.
- [19] Business process model and notation. https://fr.wikipedia.org/wiki/Business_process_model_and_notation.
- [20] Bootstrap 4. [https://fr.wikipedia.org/wiki/Bootstrap_\(framework\)](https://fr.wikipedia.org/wiki/Bootstrap_(framework)).
- [21] Jakarta ee wikipedia. https://fr.wikipedia.org/wiki/Jakarta_EE.
- [22] Hibernate ogm reference guide. <https://docs.jboss.org/hibernate/ogm/4.0/reference/en-US/htmlsingle/preface>.